基于网络评论的美食推荐系统

邓涵兮[1] 陈志华[2]
（1. 中国传媒大学国内交流与合作处，北京 100024；2. 福州大学计算机与大数据学院，福建 福州 350108）

摘 要：由于美食推荐的实时需要，本研究提出一套基于网络评论的美食推荐系统，以提供各家餐厅的介绍与评论摘要。其中，美食推荐系统主要提供功能包括有网页内容撷取机器人、多文本自动摘要技术，以自动撷取相关的评论和部落格文章，并自动提取出重要的评论句。最后，美食推荐系统结合云计算技术，为多文本自动摘要技术建立并行运算以实时提供美食评论服务。

关键词：美食推荐；信息检索；文本自动摘要；云计算　　**中图分类号：**TP391　　**文献标识码：**A
文章编号：1671-0134（2022）03-039-03　　DOI：10.19483/j.cnki.11-4653/n.2022.03.011
本文著录格式：邓涵兮，陈志华 . 基于网络评论的美食推荐系统 [J]. 中国传媒科技，2022（03）：39-41.

导语

近年来，随着人民生活质量逐渐提高，人们对于美食也越加讲究，不仅食物要满足顾客的味蕾，服务与价格也要符合顾客的期望。[1] 虽然现今网络已非常普及，人人都可以在网络上分享自己的用餐经验，然而面对众多来源的评语，要能快速且正确地认识一家餐厅仍是一件困难的事。

基于美食推荐的实时需要，文章提出一套基于网络评论的美食推荐系统“食况转播系统”，以提供各家餐厅之介绍与评论摘要。让人们可以快速决定最佳的用餐地点，甚至在陌生的环境，也能避免“踩雷”的情况发生。

1. 系统设计

本研究所设计的“食况转播系统”所提供功能包括：网页内容撷取机器人、多文本自动摘要技术（Multiple Document Summarization，MDS）[2]、云计算技术等设计。

通过网页内容撷取机器人用百度等搜索引擎对网页相关信息进行搜寻，于各个网页中找寻相关美食评论信息，撷取机器人子系统将其爬行（Crawl）数据和经过剖析（Parse）后，将相关的信息存为 Blog Corpus。最后，再利用多文本自动摘要技术，将相关网页 Corpus 中的美食评论撷取出来，并制成摘要形式，提供给用户饮食决策参考，用户可以通过本系统所设计的人机接口进行查询，整体系统处理之流程如图 1 所示。

1.1 网页内容撷取机器人

网页内容撷取机器人主要提供有模糊搜寻机制、网页爬虫（HTML Crawler），以及网页剖析器（HTML Parser）等功能，各功能说明分述如下。

1.1.1 模糊搜寻机制

模糊搜寻机制提供模糊运算与判断，建立搜寻相关的关键词字库，以关键词字库内容主动向百度搜寻进行搜寻。

1.1.2 网页爬虫

网页爬虫将百度搜寻后结果（如回传的各个网页内容）进行爬行，追踪相关连结网页并将 HTML 内容暂存。

1.1.3 网页剖析器

网页剖析器将网页爬虫取得的网页进行 HTML tag 解读，取得主要信息，并有效去除相关特殊字符（如单引号和双引号）和避免数据库隐码攻击等问题，建立 Web Corpus 以利后续之多文本自动摘要之推论。

1.2 多文本自动摘要技术

“食况转播系统”结合多文本自动摘要技术，实时将各个网页中相关美食网站的评论进行自动摘要，有效减少信息量，提取出重点评论摘要，让使用者能快速浏览过去吃过该餐厅或美食消费者的看法与经验。

多文本自动摘要技术主要参考 MEAD 套件 [3] 进行系统实践，将网页 Corpus 中相关之美食评论输入至自动摘要模块中，并由于数据庞大需有效和快速的平行运算，故将把此模块实践于 Hadoop 平台中，并以 MapReduce 进行实践，其通过数据预先处理（Preprocess）、特征选取（Feature Selected）、分类器（Classifier）、重新排序器（Reranker）、产出摘要（Summery）等步骤进行自动摘要提取，详细功能设计分述如下。

1.2.1 数据预先处理

将网页内容撷取机器人处理后的 HTML 进行撷取，并依序定义各个文章（Document）编号和语句（Sentence）编号，以进行各语句权重计算和摘要产生。

1.2.2 特征选取

基金项目：国家自然科学基金资助项目（项目编号：61906043）

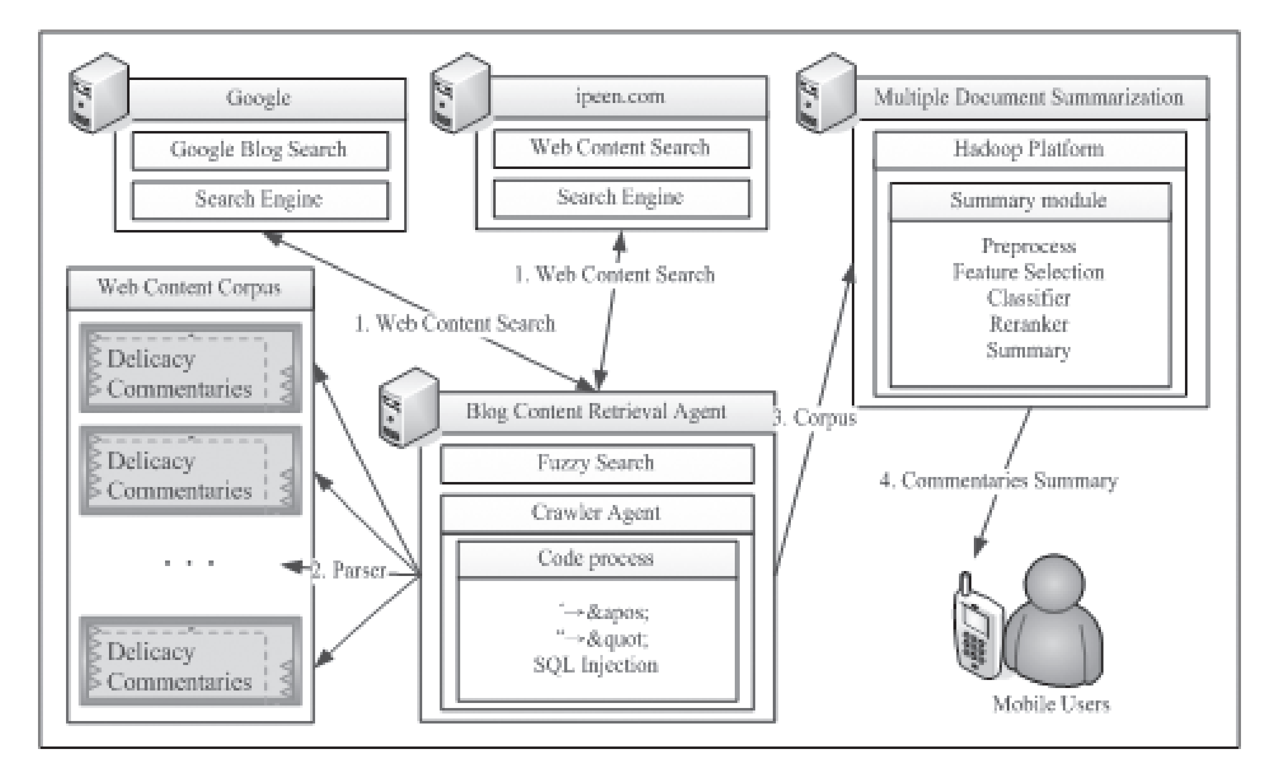


图 1 食况转播系统流程图

“食况转播系统”主要采用主题字词（Thematic Words）和评论字词（Comments Terms）两个特征（Feature）进行字词子句的权重计算。

1.2.2.1 主题字词

计算某个语句的主题字词出现的次数，当出现的次数越多则代表该语句与目标主题的关系越强烈。[4] 对于评论文件中的第 i 个语句 si 而言，该语句 si 共包含 ni 个字词 w，主题字词评分计算方式如公式（1）所示。

$$f_1(s_i)=1+\sum_{j=1}^{n_i}a_{i,j},\text{where}$$
$$a_{i,j}=\begin{cases}1,\text{word } w_{i,j} \text{ in } s_i \text{ and } w_{i,j}\in\text{主题字}\\0,\text{otherwise}\end{cases}\quad(1)$$

1.2.2.2 评论字词

计算某个语句的评论字词出现的次数，当出现的次数越多则代表该语句越具评论意义。[5] 对于评论文件中的第 i 个语句 si 而言，该语句 si 共包含 ni 个字词 w，评论字词评分计算方式如公式（2）所示。

$$f_2(s_i)=1+\sum_{j=1}^{n_i}b_{i,j},\text{where}$$
$$b_{i,j}=\begin{cases}1,\text{word } w_{i,j} \text{ in } s_i \text{ and } w_{i,j}\in\text{评论字词}\\0,\text{otherwise}\end{cases}\quad(2)$$

1.2.3 分类器

就每个特征来讨论，每个特征的重要程度有所不同，分类器主要在于做加权总和，计算出各个语句的权重，计算方式如公式（3）所示。

$$F(s_i)=f_1(s_i)\times f_2(s_i)\quad(3)$$

1.2.4 重新排序器

主要在于重新计算语句与语句之间的相似度，并设定门坎值以进行过滤，取出重要且彼此之间相似度不会太高的语句，最后再依设定的压缩率进行提取（extract）。

1.2.5 产出摘要

将重新排序器所提取出的语句顺序，依数据预先处理之文章（Document）编号、语句（Sentence）编号和原始评论文件进行对应（Mapping），取得多评论自动摘要内容，并把最后结果产出，提供给使用者快速浏览参考。

1.3 云计算技术

网络充斥着大量且繁杂的网页内容，当分析网页内容时将会因为网页数量和内文数量而造成的大量运算。由于执行效能考虑，文章将采用云计算进行平行处理，以 Hadoop 平台进行实践（Chen et al.，2012），将每篇评论文章的语句分别执行，以快速地计算每个语句的分数，并取得最重要的语句，提供使用者决策参考。

2. 系统实践

本研究设计的系统可提供给一般民众使用，使用者可以通过手机连结至“食况转播系统”，再由系统提供各家餐厅的介绍与评论摘要。让人们可以快速地决定最佳的用餐地点，甚至在陌生的环境，也能避免误”踩地

雷“的情况发生。

“食况转播系统”中，使用者端可达到各个美食餐厅的简介、各个美食的相关评论。本研究通过网页内容撷取机器人撷取相关的美食评论文章，并通过多文本自动摘要技术提供美食评论摘要，以提供使用者饮食决策参考。如图 2 所示，使用者可输入欲查询的店家名称，以搜寻该店家的相关美食评论摘要，以下以“夏慕尼”为例进行说明。当使用者输入店家名称，并点击“美食评论”时，提供该店家过去消费过的使用者经验，并进行文本自动摘要，通过算法摘录出重要的评论语句，让使用者可以快速地进行决策参考，如图 3 所示。最后，提供地图导览功能（如图 4 所示）引导消费家前往餐厅。



图 2　主功能画面

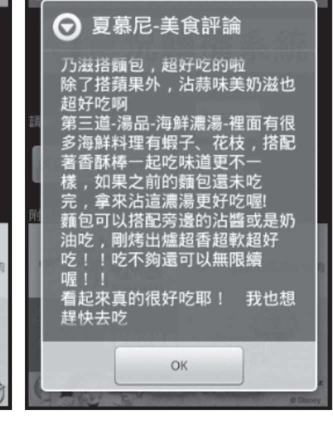


图 3　美食评论画面



图 4 地图导览画面

结语

本研究着重于使用者对餐厅选择的决策需要，发展一套基于网络评论的美食推荐系统“食况转播系统”，结合人工智能和信息检索技术，从“传媒”向“智媒”转变[6-7]，搜集并统计餐厅相关信息的推荐，并结合餐厅介绍与相关评论，将网络信息（例如：博客、爱评网、以及 Blog 等相关美食评论）进行自动摘要处理，供使用者快速认识该餐厅，评估是否合适作为用餐地点。未来可以尝试将此系统模型应用于各行各业的评论摘要和推荐信息中，例如旅游业。媒

参考文献

[1] 周蕾，李强．基于 LBS 应用的淮安美食推荐类系统的研究 [J]. 食品安全导刊，2021（21）：172–173.

[2] 王青，松张衡，李菲．基于文本多维度特征的自动摘要生成方法 [J]. 计算机工程，2020（9）：110–116.

[3]Hui–Fei Lin，Chi–Hua Chen，J.M. An Intelligent Embedded Marketing Service System Based on TV Apps：Design and Implementation through Product Placement in Idol Dramas [J]. Expert Systems with Applications，2013（10）：4127–4136.

[4] 罗芳，汪竞航，何道森，蒲秋梅．融合主题特征的文本自动摘要方法研究 [J]. 计算机应用研究，2021（1）：129–133.

[5] 章成志，童甜甜，周清清．基于细粒度评论挖掘的书评自动摘要研究 [J]. 情报学报，2021（2）：163–172.

[6] 徐曼．从“传媒”向“智媒”转变——人工智能技术对新闻业产生的影响 [J]. 中国传媒科技，2021（7）：56–58.

[7] 吴周可心．智能媒介技术与商业传播的互构与互驯 [J]. 中国传媒科技，2021（8）：103–105.

作者简介：邓涵兮（1986–），女，北京，讲师，研究方向：文化产业、影视传媒；陈志华（1984–），男，中国台湾，教授，研究方向：人工智能、物联网。

（责任编辑：胡杨）